

Open Data im hbz-Verbund

Adrian Pohl, Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen

<pohl@hbz-nrw.de>

Mit dem Internet und – als dessen wichtigstem Bestandteil – dem World Wide Web formt sich seit einigen Jahrzehnten eine umfassende Publikations- und Kommunikationsplattform aus, auf der zukünftig der Großteil aller Publikation und Kommunikation stattfinden wird. Als eine Erweiterung des bestehenden Webs lässt sich Linked Open Data verstehen. Mit Linked Open Data werden zwei Standards bezeichnet, die die Funktionalität eines Netzes von Daten sichern sollen, indem sie die rechtliche und technische Kompatibilität von Daten im Web garantieren:

- *Open-Data-Standards* sorgen für die *rechtliche* Basis der Nutzung und Kombination verteilter Daten im Netz.
- *Linked-Data-Standards* sorgen für die *technische* Kompatibilität zwischen verteilt vorliegenden Daten.

In einer dreiteiligen Artikelreihe über Linked-Open-Data-Aktivitäten am Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (hbz) sollen die rechtlichen wie technischen Dimension von Linked Open Data erläutert werden und die Notwendigkeit, die Ziele und der Nutzen von Linked Open *Bibliographic Data*¹ dargelegt werden.

Im ersten Teil dieser Reihe über das Was, Warum und Wie von Linked-Open-Data-Aktivitäten am hbz sollen einige Fragen zu *Open Data* geklärt werden. Er erscheint gedruckt in ProLibris 3/2010. Der zweite Teil – gemeinsam verfasst von Felix Ostrowski und Adrian Pohl – mit dem Schwerpunkt *Linked Data* erscheint gedruckt in B.I.T. online 3/2010 und der dritte, in dem sich ebenfalls Felix Ostrowski und Adrian Pohl mit der Konzeption und Implementierung einer Open-Data-Infrastruktur befassen, wird gedruckt im Tagungsband der DGI-Konferenz *Semantic Web & Linked Data Elemente zukünftiger Informationsinfrastrukturen* publiziert.

Alle Texte werden darüber hinaus unter einer CC-BY-Lizenz im Web publiziert, siehe etwa unter <http://www.hbz-nrw.de/dokumentencenter/produkte/lod/>.

¹ Die Ausdrücke „Freie Katalogdaten“ und „Open Bibliographic Data“ werden in diesem Kontext synonym verwendet.

1 Einleitung

Nachdem im Januar die CERN Library² und im Februar die Universitätsbibliothek Ghent³ die Freigabe ihrer Katalogdaten verkündet hatten, verbreitete sich im März dieses Jahres in bibliothekarischen Kreisen und unter Open-Data-Interessierten weltweit eine Nachricht über Mailinglisten, Twitter und Blogs: Kölner Bibliotheken und das Landesbibliothekszenrum Rheinland-Pfalz hatten in Kooperation mit dem hbz ihre Katalogrohdaten der Öffentlichkeit als Ganzes unter einer Public-Domain-Lizenz zur freien Verfügung gestellt.⁴ Dazu wurden die Daten aus dem Verbundkatalog⁵ bzw. aus den Lokalsystemen⁶ exportiert, beschrieben, auf einem Webserver deponiert und mit einer Public-Domain-Lizenz (siehe Abschnitt 4) versehen. Diese Daten sind nun *Open Data*. Das hbz arbeitet daran, sie in *Linked Data* zu überführen, damit sie Teil des Semantic Web werden.⁷

Im hbz war der ausschlaggebende Punkt für eine Datenfreigabe letztlich das Ziel, die Verbunddaten ins Semantic Web zu bringen.⁸ Dieses Ziel entstand im hbz im Jahre 2009, und nach einer tiefergehenden Auseinandersetzung mit den technischen, politischen und rechtlichen Aspekten von Linked Data kamen die Beteiligten zu der Einsicht, dass die Lizenzfrage absolute Priorität hat. Im hbz galt also ziemlich bald schon die Losung: „Erst öffnen, dann linken!“⁹

Aber was bedeutet ‚Open Data‘ genau? Und was ist Linked Data? Warum beschäftigen wir uns damit und welche Entwicklungen finden in diese Richtung statt? In diesem ersten Teil sollen einige Fragen zu Open Data beantwortet werden.

² Siehe die Pressemitteilung CERN (2010): The CERN Library publishes its book catalog as Open Data. Einsehbar unter <http://library.web.cern.ch/library/Library/announcement.html>.

³ <http://lib.ugent.be/info/en/exports.shtml>, letzter Zugriff am 15.7.2010.

⁴ Siehe die Pressemitteilung Hochschulbibliothekszenrum des Landes Nordrhein-Westfalen (2010): Freigabe der Katalogdaten: Kölner Bibliotheken leisten Pionierarbeit. Einsehbar unter: <http://www.hbz-nrw.de/dokumentencenter/presse/pm/datenfreigabe>. Eine Sammlung der Reaktionen findet sich in einem Blogbeitrag und den dazugehörigen Kommentaren, siehe Flimm, Oliver (2010): Offene Katalogdaten – Reaktionen. Einsehbar unter <http://blog.openbib.org/2010/03/18/offene-katalogdaten-reaktionen/>. Der Autor des vorliegenden Textes hat zum Thema bereits einige Blogbeiträge veröffentlicht, die als Basis dieses Textes dienen, siehe <http://www.uebertext.org/search/label/opendata>.

⁵ Informationen und Downloadlinks zu den Exporten aus dem hbz-Verbundkatalog finden sich hier: <http://opendata.hbz-nrw.de/projects/data-publishing/wiki/Download-de>.

⁶ Die Universitäts- und Stadtbibliothek Köln (<http://opendata.ub.uni-koeln.de/>) wie auch die Zentralbibliothek der Sportwissenschaften (<http://opendata.zbsport.de/>) der Deutschen Sporthochschule Köln haben zusätzlich die Daten aus ihren Lokalsystemen freigegeben, weil diese zusätzliche Informationen enthalten und so die Möglichkeit besteht, in anderen Formaten zu publizieren.

⁷ Vgl. dazu den zweiten Teil dieser Artikelreihe Ostrowski, Felix / Pohl, Adrian (2010): Linked Data im hbz. Erscheint in B.I.T. Online 3/2010 und online u.a. unter <http://www.hbz-nrw.de/dokumentencenter/produkte/lod/>.

⁸ Dies gilt nicht für alle Kooperationspartner. Die Universitäts- und Stadtbibliothek Köln betonte immer auch den Nutzen von Open Data im allgemeinen auch ohne die Umwandlung der Daten in Linked Data. Zweifelsohne besteht dieser Nutzen auch, siehe Abschnitt 6.

⁹ Vgl. den Blogbeitrag Pohl, Adrian (2009): Erst öffnen, dann linken!. Einsehbar unter <http://www.uebertext.org/2009/11/erst-offnen-dann-linken.html>.

2 Open Data – Historisches

Seit etwa 25 Jahren existiert die Open-Source-Bewegung für freie Software¹⁰ und die Open-Access-Bewegung hat ihre Wurzeln in den 90er Jahren des letzten Jahrhunderts. Der Begriff *Open Data* hingegen ist noch vergleichsweise jung.¹¹ Dahinter stecken die Bemühungen zur Etablierung eines Konzept und einer Praxis der Offenheit und Freiheit auch im Umgang mit Rohdaten.

Vor allem die Aktivitäten in den USA und in Großbritannien im Bereich öffentlicher administrativer Daten (Open Government Data)¹² haben dem Open-Data-Gedanken in der letzten Zeit große öffentliche Aufmerksamkeit eingebracht.¹³ Aber auch in den Wissenschaften wird der dieser Gedanke vorangetrieben, was etwa Projekte wie Science Commons¹⁴ oder die Verabschiedung der Panton Principles¹⁵ zeigen.

In der Bibliothekswelt kam das Thema Open Data bereits im Jahr 2007 auf: zum einen mit dem Start der Open Library¹⁶ und zum anderen mit einer Diskussion über den Berichtsentwurf *Future of Bibliographic Control*¹⁷ der Library of Congress (LoC). Dieser Bericht ließ die Fragen der Lizenzierung von bibliographischen Daten außen vor und zog dadurch Kritik auf sich.¹⁸

¹⁰ Diese Bewegung hat schon vor längerer Zeit begonnen, einen klaren Begriff von „Free Software“ herauszuarbeiten, siehe Free Software Foundation (2010): The Free Software Definition. Einsehbar unter <http://www.gnu.org/philosophy/free-sw.html> oder Open Source Initiative: Open Source Definition. Einsehbar unter <http://www.opensource.org/docs/osd>.

¹¹ Zur Geschichte des Begriffs siehe auch den Blogbeitrag Murray-Rust, Peter (2006): Open Data - the time has come. Einsehbar unter <http://wwwmm.ch.cam.ac.uk/blogs/murrayrust/?p=32>.

¹² Siehe die Webseiten <http://data.gov> und <http://data.gov.uk>, auf denen Daten öffentlicher Institutionen, die nicht personenbezogen oder sicherheitsrelevant sind, der Öffentlichkeit im vollen Umfang zur Verfügung gestellt werden.

¹³ In Deutschland hinkt die Entwicklung im Vergleich zu Großbritannien und den USA um einige Jahre hinterher. Allerdings gibt es erste Initiativen, die sich für Open Government Data in Deutschland einsetzen, z.B. das Open-Data-Network (<http://opendata-network.org/network/>), in dem auch Vertreter aus der Politik aktiv sind, und der deutsche Zweig der Open Knowledge Foundation (<http://okfn.de/>).

¹⁴ <http://sciencecommons.org/>

¹⁵ Diese Panton Principles sprechen sich für Open Data in der Wissenschaft aus, damit eine optimale Basis für wissenschaftliches Arbeiten gefördert und die Teilhabe der Gesellschaft an den Ergebnissen der Wissenschaft ermöglicht wird.

¹⁶ <http://openlibrary.org>. Siehe auch den Artikel „Open Library“ in: Wikipedia, Die freie Enzyklopädie.

Bearbeitungsstand: 9. Mai 2010, 02:20 UTC. URL:

http://de.wikipedia.org/w/index.php?title=Open_Library&oldid=74131950.

¹⁷ Library of Congress Working Group on the Future of Bibliographic Control (2007): Report on the Future of Bibliographic Control. Draft for Public Comment. Einsehbar unter <http://www.loc.gov/bibliographic-future/news/lcwg-report-draft-11-30-07-final.pdf>.

¹⁸ In einer auf Initiative der Open Knowledge Foundation veröffentlichten Antwort auf diesen Bericht wurde die LoC aufgefordert, sich klar für eine offene Lizenzierung von Katalogdaten auszusprechen. Siehe Open Knowledge Foundation (2007): Response to Working Group on the Future of Bibliographic Control (Library of Congress). Einsehbar unter <http://wiki.okfn.org/FutureOfBibliographicControl>.

3 Was ist Open Data?

Wann kann man bestimmte Daten berechtigterweise als Open Data bezeichnen? Sobald sie frei im Internet zugänglich sind? Wieso sind Katalogdaten nicht schon Open Data, wenn doch jeder die Daten in Online-Bibliothekskatalogen recherchieren kann?

Der Ausdruck „open“ wird in vielerlei Kontexten und mit verschiedensten Bedeutungen verwendet. Deshalb ist es, sinnvoll, diesen Begriff zunächst klarer zu umreißen. Orientierung gibt die *Open Definition*¹⁹, ein Projekt der Open Knowledge Foundation (OKFN)²⁰.



Abbildung 1: Arten von Wissen gemäß der Open Definition

Die Open Definition dient dem Zweck, das Konzept der Offenheit im Kontext von *Open Knowledge* zu spezifizieren. Der allgemeine Begriff ‚Open Knowledge‘ umfasst Inhalte (Texte, Bücher, Filme), Daten sowie Informationen aus Regierung und öffentlicher Verwaltung (vgl. Abb. 1).²¹ Das Ziel der Open Definition ist es, eine gewisse Verbindlichkeit des Redens über Open Knowledge zu erreichen. Der wichtigste Aspekt der Open Definition ist hier, dass sie die rechtliche Kompatibilität verteilter Wissensressourcen im Web ermöglicht. Das bedeutet: Alle Inhalte oder Daten, die mit einer Open-Definition-konformen Lizenz versehen sind, können ohne weitere Klärung von Rechtsfragen abgefragt, kopiert, verteilt, kombiniert und erweitert werden.

Die Kurzform dieser Definition von "offen" besagt:

"A piece of knowledge is open if you are free to use, reuse, and redistribute it."²²

Die vollständige Definition spezifiziert die Zugangsbedingungen zu offenem Wissen wie folgt: *The work shall be available as a whole and at no more than a reasonable reproduction cost, preferably downloading via the Internet without charge.*²³ Offen ist also ein Bibliothekskatalog im Sinne der Open Definition nur, wenn die Daten auch als Gesamtheit heruntergeladen und weiterverwendet werden können. Klassische Bibliotheks- und Verbundkataloge sind demnach nicht offen.

¹⁹ Open Knowledge Foundation: Open Definition. Einsehbar unter <http://www.opendefinition.org>.

²⁰ <http://okfn.org>. Zur Mitarbeit des hbz in der OKFN siehe Abschnitt 5.

²¹ Software lässt die Open Definition bewusst außen vor, weil "Open Source Software"/„Free Software“ eben schon klar definiert ist, vgl. Fußnote 9.

²² Siehe <http://www.opendefinition.org/>.

²³ <http://opendefinition.org/1.0/>.

4 Open-Data-Lizenzierung

Die Frage der Lizenzierung von Internetinhalten fand in der Vergangenheit in der Bibliothekswelt (wie in vielen anderen Bereichen auch) zu wenig Beachtung. Mittlerweile beginnt sich allerdings in den Gedächtnisinstitutionen ein Bewusstsein für die Bedeutsamkeit klarer menschen- wie maschinenlesbarer Nutzungsbedingungen im Web zu entwickeln. Die Grundregel für jedwede Veröffentlichung von Inhalten, Daten, künstlerischen Artefakten etc. im Internet lautet:

Webinhalte sollten mit leicht verständlichen menschen- und maschinenlesbaren Lizenzinformationen versehen werden.

Diese Aussage gilt ganz allgemein, unabhängig davon, ob jemand Nachnutzungen (z.B. Wiederveröffentlichungen) verbieten möchte oder Offenheit als Ziel verfolgt. Ob es nun heißt „Alle Rechte vorbehalten“, „Einige Rechte vorbehalten“²⁴ oder „Alles ist erlaubt“, wichtig ist, dass diese Angaben zunächst einmal überhaupt vorhanden sind: Menschen und bestenfalls auch Maschinen sollten im Netz problemlos erkennen können, was ihnen im Umgang mit bestimmten Web-Inhalten rechtlich erlaubt ist, ohne sich mit den Urhebern in Verbindung setzen zu müssen.

Um etwas als offene Daten zu publizieren, müssen bestimmte Nutzungsbedingungen angegeben werden. Damit Daten offen sind, d.h. damit sie gemäß der Open Definition „Open Data“ sind, müssen sie mit einer Lizenz versehen sein, die den Kriterien der Offenheit genügt.²⁵ Es gibt hier drei Möglichkeiten:

- Eine *Public-Domain-Lizenz* gibt allen die Möglichkeit, die Daten zu kopieren und ohne jegliche Einschränkung nach Belieben zu verwenden.²⁶
- Eine *Attribution-Lizenz* gibt allen die Möglichkeit, die Daten zu kopieren nach Belieben zu verwenden, allerdings mit der Einschränkung, dass bei Nachnutzung der Daten auf den Urheber bzw. die Ursprungsdatenbank verwiesen werden muss.²⁷
- Eine *Share-Alike-Lizenz* gibt allen die Möglichkeit, die Daten zu kopieren und nach Belieben zu verwenden, allerdings mit den Einschränkungen, dass 1.) der Urheber der verwendeten Daten genannt werden muss und 2.) die aus der Nachnutzung entstehenden Daten oder Werke unter einer gleichwertigen Lizenz veröffentlicht werden müssen.²⁸

Wohlgemerkt: Sogenannte Non-Commercial-Lizenzen (NC-Lizenzen), die eine kommerzielle Nutzung der freigegebenen Daten oder anderer Inhalte verbieten, sind nicht mit offen im Sinne der Open Definition und sollten aus mindestens zwei Gründen nicht verwendet werden:

²⁴ Für die Lizenzierung von Webinhalten gibt es seit einigen Jahren die Creative-Commons-Lizenzen, die nach dem „Some rights reserved“-Prinzip vorgehen und Urhebern die Möglichkeit bieten, einige Nachnutzungen ihrer Werke zu erlauben, während andere verboten werden: <http://creativecommons.org/>.

²⁵ Eine Übersicht über die Open-Definition-konformen Datenlizenzen, an die sich die folgende Übersicht anlehnt, gibt es hier: <http://opendefinition.org/licenses#Data>.

²⁶ Public-Domain-Lizenzen für Daten sind die CCZero (<http://creativecommons.org/publicdomain/zero/1.0/>) und die PDDL (<http://www.opendatacommons.org/licenses/pddl/>).

²⁷ Eine Attribution-Lizenz für Daten ist die Open Data Commons Attribution License (<http://www.opendatacommons.org/licenses/by/>).

²⁸ Die Open Database License (ODbL) ist eine Share-Alike-Lizenz: <http://www.opendatacommons.org/licenses/odbl/>.

1. Die *Unschärfe* der Lizenz: Es ist unklar, wann ein Angebot kommerziell ist und wann nicht und welche Art der Nutzung (einzelner Daten, eine Abfrage, ein Link auf die Daten?) verboten ist. Mit einer NC-Lizenz erreicht man im Einzelfall also nicht den beabsichtigten Zweck, Klarheit über Nutzungsbedingungen herzustellen.
2. Die *Inkompatibilität* der Lizenz: Non-Commercial-Lizenzen verhindern die Kombination verschieden lizenzierter Inhalte. Zum Beispiel sind Non-Commercial-Lizenzen inkompatibel mit Share-Alike-Lizenzen, weil diese nämlich eine kommerzielle Nutzung erlauben. Das heißt etwa, dass die Wikipedia, deren Inhalte unter einer Share-Alike-Lizenz publiziert sind, mit einer NC-Lizenz versehene Katalogdaten nicht nachnutzen kann.²⁹

Das hbz und die kooperierenden Bibliotheken haben sich für die Freigabe der Katalogdaten auf die freieste Lizenz geeinigt, eine Public-Domain-Lizenz (CCZero).

5 Offene Standards

Die Open Definition beschränkt sich allerdings nicht auf den lizenzrechtlichen Aspekt, sondern es werden noch einige andere Anforderungen an Open Knowledge gestellt. Hier soll noch eine weitere Anforderung näher betrachtet werden: das Erfordernis offener Standards. Damit Open Data anderen einen Nutzen bringen kann, müssen freigegebene Daten verstanden werden können, d.h. Speicherformate, Datenmodelle, Metadatenelemente usw. müssen an irgendeiner Stelle dokumentiert sein. Artikel 4 der Open Definition lautet entsprechend:

*The work must be provided in such a form that there are no technological obstacles to the performance of the above activities. This can be achieved by the provision of the work in an open data format, i.e. one whose specification is publicly and freely available and which places no restrictions monetary or otherwise upon its use.*³⁰

Bibliotheksdaten liegen zwar in offenen Standards vor, weil es offen zugängliche MAB- und MARC-Spezifikationen gibt. Allerdings sind diese Standards für Nichtbibliothekare schlichtweg zu undurchsichtig und zu schwer verständlich, so dass sie faktisch eine Nachnutzung durch Akteure außerhalb der Bibliothekswelt verhindern.

Das hbz, die Universitäts- und Stadtbibliothek Köln wie auch die Zentralbibliothek der Sportwissenschaften der Deutschen Sporthochschule Köln haben ihre Daten in auf MAB-basierten-Formaten herausgegeben, die zwar im Netz dokumentiert sind, allerdings ausschließlich auf Deutsch. Das Ergebnis ist, dass nur MAB-geschulte Menschen aus dem deutschsprachigen Raum überhaupt damit etwas anfangen können und dass die gesamte englischsprachige Welt die Freigabe zwar registriert hat aber von einer Nachnutzung aus Gründen der beschriebenen Schwierigkeiten Abstand nimmt.

Diese Umstände unterstreichen wie wichtig es ist, bibliographische Daten in webkompatible, international anerkannte Standards zu überführen, damit sich das volle Potential von Open Data überhaupt entfalten kann. Die genannten Schwierigkeiten sollen aber niemandem als Ausrede

²⁹ Für eine ausführliche Kritik von Non-Commercial Lizenzen vgl. den „Klassiker“ zum Thema: Möller, Erik (2005): Freiheit mit Fallstricken: Die Creative Commons NC-Lizenzen und ihre Folgen. Einsehbar unter <http://freedomdefined.org/Licenses/NC/De>.

³⁰ <http://www.opendefinition.org/okd/>

dienen, Katalogdaten gar nicht erst freizugeben, denn auch schwer nachnutzbare frei Daten können nützlich sein, wie im nächsten Abschnitt beleuchtet wird.

Offene Standards als Kriterium für Open Data markieren gewissermaßen den Punkt, an dem sich die Konzepte *Open Data* und *Linked Data* berühren. Hier mag auch die Ursache dafür liegen, dass die beiden Begriffe häufig synonym und austauschbar verwendet werden. Eine klare Unterscheidung scheint aber dennoch sinnvoll und nützlich, weshalb eine solche auch in dieser Artikelreihe getroffen wird.

6 Warum Open Data?

Es gibt eine ganze Reihe von Gründen, die für die Freigabe von Katalogdaten sprechen. Sowohl für einzelne Bibliotheken und Verbünde als auch für die Bibliothekswelt als Ganzes und natürlich für alle, die bibliographische Daten nachnutzen wollen und können bringt eine Datenfreigabe Vorteile. Im Folgenden werden einige genannt.³¹

Konsequente Umsetzung des Open-Access-Gedankens. In der *Berliner Erklärung über den offenen Zugang zu wissenschaftlichem Wissen* von 2003 heißt es:

*Die Vision von einer umfassenden und frei zugänglichen Repräsentation des Wissens lässt sich nur realisieren, wenn sich das Internet der Zukunft durch Nachhaltigkeit, Interaktivität und Transparenz auszeichnet. Inhalte und Software müssen offen zugänglich und kompatibel sein.*³²

Bibliotheken und Bibliothekar/innen fühlen sich dem Open-Access-Gedanken verpflichtet und fördern die Open-Access-Bewegung seit langem. In den Daten der Bibliothekskataloge ist eine Menge Wissen über die Welt der (wissenschaftlichen) Literatur enthalten, welches ebenfalls „offen zugänglich und kompatibel“ bereitgestellt werden sollte.

Nutzen maximieren & Sichtbarkeit steigern, Partizipation & neue Dienste ermöglichen. Ganz allgemein bringt die Freigabe von Katalogdaten Bibliotheken und Verbünde vier Hauptzielen näher:

- Sicher zu stellen, dass der Reichtum bibliographischer Daten maximalen Nutzen bringt
- Die Sichtbarkeit bibliothekarischer Institutionen im Web zu steigern und dadurch ihre Relevanz im Internetzeitalter zu sichern
- Die Gruppe der Akteure bei der Erstellung und Verbesserung bibliographischer Daten zu erweitern
- Die Entwicklung einer völlig neuen Generation von Diensten und Anwendungen für Forscher/innen und Bibliotheksbenutzer/innen zu ermöglichen

³¹ Der folgende Text basiert auf dem Entwurf eines Open-Bibliographic-Data-Flyers, dessen ursprüngliche Version im hzb entstanden ist: Open Knowledge Foundation (2010): Entwurf eines Flyers zu Freien Katalogdaten. Einsehbar unter <http://wiki.okfn.org/Entwurf%20eines%20Flyers%20zu%20Freien%20Katalogdaten?action=recall&rev=9> sowie auf der in der Working Group on Open Bibliographic Data geleisteten Überarbeitung der englischen Übersetzung (siehe <http://okfnpad.org/ep/pad/view/bibliography-flyer/FnGzGN6MWI>).

³² Berliner Erklärung über den offenen Zugang zu wissenschaftlichem Wissen (2003). Deutsche Version einsehbar unter http://oa.mpg.de/openaccess-berlin/Berliner_Erklaerung_dt_Version_07-2006.pdf.

Vorteile für Benutzer/innen. Abgesehen von der Optimierung der Rechercheinstrumente geben freie Katalogdaten allen Interessierten die Möglichkeit, bibliographische Daten einfacher in ihre eigenen Arbeiten zu integrieren, etwa durch

- die leichtere Erstellung von Bibliographien,
- die bessere Integration der Daten in Lern- und Lehrumgebungen und
- die zunehmende Verzahnung der bibliographischen Daten in Wissenschaft und Bibliotheken.

Optimierung der Kooperation zwischen bibliothekarischen Institutionen. Kooperation ist seit langem eine feste Größe innerhalb der Bibliothekswelt.³³ Eine weit verbreitete Open-Bibliographic-Data-Praxis würde die Bedingungen dieser Kooperation verbessern, da die bestehenden nationalen Grenzen wie auch Verbundgrenzen stärker überschritten und letztlich aufgelöst werden könnten. Neue Möglichkeiten der Fremddatenübernahme und Kataloganreicherung können durch Open Data entstehen. Es wird z.B. möglich, eigene Kataloge mittels der bereitgestellten Daten anderer Bibliotheken mit Informationen, z.B. Schlagwörtern und Klassifikationsstellen anzureichern.

Nutzbarkeit durch nichtbibliothekarische Institutionen. Auch eine große Anzahl nichtbibliothekarischer Institutionen, wie beispielsweise die Wikipedia³⁴ oder das Internet Archive³⁵, sind sehr interessiert, Daten aus der Bibliothekswelt in ihre Dienste zu integrieren. Dies eröffnet zudem eine Perspektive zur Kooperation mit solchen Organisationen und damit zu einer stärkeren Integration der Bibliothekswelt ins Web.

Feedback und Datenaufwertung durch andere. Offene Daten bieten anderen die Möglichkeit Feedback zu geben, durch dessen Berücksichtigung die Datenqualität verbessert werden kann.³⁶ Mit den Webinhalten entstehen immer größere Mengen auch wissenschaftlich relevanter Informationen, die durch Bibliothekare allein nicht mehr gemanagt werden können. Es ist an der Zeit, über ein „Crowdsourcing“ der Datenpflege und -anreicherung nachzudenken und die Basis der Mit"arbeiter" um eine theoretisch unbegrenzte Menge Freiwilliger zu erweitern.³⁷

Die Rückführung von durch die Arbeit anderer Akteure aufgewerteten Daten in die Bibliothekskataloge könnte eine Folge der Kooperation sein und erheblichen Nutzen bringen. Katalogdaten eignen sich aufgrund ihrer hohen Qualität und der vertrauenswürdigen Herkunft etwa auch als Grundlage wissenschaftlicher Forschungen (z.B. in den Literaturwissenschaften) und könnten dadurch letztlich auf Arten und Weisen aufbereitet werden, an die wir selber momentan

³³ Spätestens mit dem *System Althoff* entstanden um die Wende vom 19. zum 20. Jahrhundert in Preußen und darüber hinaus erste Ansätze der kooperativen Katalogisierung im Kontext eines Systems des kooperativen Bestandsaufbaus.

³⁴ <http://wikipedia.org>

³⁵ <http://archive.org>

³⁶ Dies ist im kleinen Maßstab bereits passiert als ein Nutzer Fehler in den Personendaten zu seiner Person in einem Wikimedia-Projekt fand, das die USB-Daten integrierte. Wikimedia e.V. leitete den Fehler weiter und er konnte im Katalog der USB korrigiert werden. Damit eine solche Kooperation im großen Maßstab stattfinden kann, müssten natürlich noch entsprechende Anwendungen entwickelt werden.

³⁷ Zum Crowdsourcing-Konzept vgl. den Artikel „Crowdsourcing“ in: Wikipedia, Die freie Enzyklopädie.

Bearbeitungsstand: 23. Juni 2010, 07:22 UTC. URL:

<http://de.wikipedia.org/w/index.php?title=Crowdsourcing&oldid=75903904>.

noch gar nicht denken.³⁸ So könnten neue Daten auf der Basis von Katalogdaten entstehen, derer sich Bibliotheken wiederum wie aus einem Fremddatenpool bedienen können, um ihre eigenen Angebote zu verbessern. Dadurch wird der Wert der bibliographischen Daten und der bibliothekarischen Angebote erhöht: Je mehr die Daten integriert, verlinkt und benutzt werden, desto umfangreicher und wertvoller werden sie letztlich.

Das Semantic Web mitweben. Letztendlich schafft die Freigabe bibliographischer Daten die nötige rechtliche Grundlage, um bibliographische Daten in einem kooperativen, offenen und transparenten Prozess in *Linked Data* zu überführen und in das Semantic Web zu migrieren. Tim Berners-Lee, der Erfinder des World Wide Web und des Semantic Web, hat die Notwendigkeit offener Rohdaten für den Erfolg des Linked-Data-Projekts in die Parole gefasst: „Raw data now!“³⁹ Mit Linked Data würden Bibliothekskataloge, welche die Daten zurzeit nur begrenzt zugänglich machen, endlich zu einem integralen Bestandteil des WWW werden.

7 Open-Data-Aktivitäten im hbz-Verbund

Von Anfang an war im hbz-Verbund klar, dass das alleinige Publizieren von Katalogdaten auf einem Webserver nur der erste Schritt zur Etablierung einer Open-Bibliographic-Data-Praxis sein kann. Auf dem eingeschlagenen Weg sind noch eine Menge Fragen technischer, politischer und kultureller Art zu beantworten.

Vernetzung und Kooperation

Da die Open-Bibliographic-Data-Bewegung noch am Anfang steht, ist eine Vernetzung und Kooperation der Akteure auf dem Gebiet unerlässlich. Seit etwa einem halben Jahr gibt es im hbz in unregelmäßigen Abständen formlose Treffen einer Linked-Open-Bibliographic-Data-Arbeitsgruppe. Ursprünglich zur Vernetzung der Akteure im hbz und in der Region gedacht, ist diese Gruppe nun offen für alle, die an einer Publikation ihrer Daten als Open Data interessiert sind. Bei den Treffen findet ein reger Erfahrungsaustausch über die letzten Entwicklungen statt und die nächsten Schritte werden diskutiert. Alle Bibliotheken und Verbünde, die ihre Katalogdaten freigeben und in Linked Data überführen wollen, sind eingeladen Vertreter/innen zu entsenden.

Auf nationaler Ebene tauscht sich das hbz u.a. mit der Deutschen Nationalbibliothek aber auch mit nichtbibliothekarischen Akteuren wie Wikimedia e.V. aus.

Auf internationaler Ebene vernetzt sich das hbz vor allem über die bereits mehrfach erwähnte *Working Group on Open Bibliographic Data* der Open Knowledge Foundation.⁴⁰ Im Juni 2010 hat der

³⁸ Literaturwissenschaftler interessieren sich etwa für Beziehungen zwischen Büchern (z.B. Parodien und Stilmachungen) oder die historische Entwicklung der Relevanz bestimmter Publikationsorte. Da die Geisteswissenschaften zunehmend auch datenorientiert arbeiten (Stichwort: Digital Humanities) gibt es hier sehr viel Kooperationspotential.

³⁹ Dies tat er in einem sehr bekannten TED-Talk im Juni 2009, siehe TED (2009): Tim Berners-Lee on the next Web. Einsehbar unter http://www.ted.com/talks/tim_berniers_lee_on_the_next_web.html. Ein Jahr später zeigte er in einem weiteren TED-Talk die ersten Ergebnisse der Open-Data-Initiative, siehe TED (2010): Tim Berners-Lee: The year open data went worldwide. Einsehbar unter http://www.ted.com/talks/lang/eng/tim_berniers_lee_the_year_open_data_went_worldwide.html.

⁴⁰ Für mehr Informationen siehe die Wiki-Seite: <http://wiki.okfn.org/wg/bibliography>. Auch diese Gruppe ist offen für neue Mitstreiter/innen.

Autor des vorliegenden Textes die Koordination dieser Gruppe übernommen. Auch bestehen Kontakte zu relevanten Arbeitsgruppen am W3C, dem World Wide Web Consortium. Im Zusammenhang mit Open Data und der Konzeption einer Open-Data-Plattform ist etwa das dcat-Vokabular sehr interessant, das derzeit im *Data Catalog Vocabulary project* am W3C entwickelt wird. Der Austausch mit dieser Gruppe wird gepflegt.⁴¹

Aufklärung und Information

Ebenso wichtig wie Vernetzung und Kommunikation sind Aufklärung und Information: über die technischen und rechtlichen Hintergründe, die Vorteile und Ziele von Linked Open Data. Neben Veröffentlichungen wie dieser arbeitet das hbz an der Erstellung eines Open-Data-Flyers mit, der internationale Verwendung finden soll. Darüber hinaus soll in den nächsten Monaten gemeinsam mit dem Urheberrechtsjuristen Till Kreutzer ein rechtlicher Leitfaden zur Freigabe von Katalogdaten entwickelt und veröffentlicht werden.⁴²

Konzeption einer Open-Data-Plattform

Die Publikation und Lizenzierung von Katalogdaten reicht allein nicht aus, wenn diese Daten möglichst Vielen Nutzen bringen sollen. Im hbz wurde damit begonnen, eine technische Open-Data-Infrastruktur zu konzipieren, welche die Entwicklung und den Nutzen einer weitverbreiteten Open-Data-Praxis unterstützen soll. Diese Plattform sollte folgende Funktionalitäten unterstützen:

- *Beschreibung* der Datenpakete aus unterschiedlichen Quellen und in unterschiedlichen Formaten,
- *Speicherung* großer Datenpakete als Option,
- *Aktualisierung* jener Datenpakete, die gepflegt werden und somit kontinuierlicher Veränderung unterliegen,
- *Versionierung* der Änderungen von Datenpaketen,
- *Verlinkung* von Datenpaketen, um Beziehungen zwischen ihnen auszudrücken sowie
- *Download-Tracking*, wenn auch die Daten und nicht nur die Beschreibungen gespeichert sind.

Die Aktivitäten in diesem Projekt werden im dritten Teil dieser Artikelreihe näher erläutert.⁴³ Darüber hinaus sind Linked-Data-Aktivitäten am hbz, bei denen eine Freigabe der zugehörigen Daten noch nicht stattgefunden hat, Thema im zweiten Teil.

⁴¹ Vgl. hierzu den dritten Teil dieser Reihe.

⁴² Für diesen Leitfaden ist Input erwünscht: Es geht darum, konkrete rechtliche Fragen und Probleme zu beleuchten, die die Freigabe eines Bibliotheks- oder Verbundkatalogs mit sich bringt. Alle an einer Datenfreigabe interessierten Bibliotheken, die vor rechtlichen Problemen und Unklarheiten stehen, sind eingeladen, sich mit dem Autor dieses Beitrags in Verbindung zu setzen.

⁴³ Dieser dritte Teil mit dem Titel „Eine Plattform für freie bibliographische Daten – Konzeption und Implementierung“ wird im Tagungsband der DGI-Konferenz *Semantic Web & Linked Data Elemente zukünftiger Informationsinfrastrukturen* veröffentlicht werden sowie – wie die anderen Teile auch – Open Access im Netz zur Verfügung stehen, u.a. anderem unter <http://www.hbz-nrw.de/dokumentencenter/produkte/lod/>.

8 Quellen

Wo nicht anders angegeben gilt für alle Webquellen: Letzter Abruf am 15.7.2010.

Berliner Erklärung über den offenen Zugang zu wissenschaftlichem Wissen (2003). Deutsche Version einsehbar unter http://oa.mpg.de/openaccess-berlin/Berliner_Erklaerung_dt_Version_07-2006.pdf.

CERN Library (2010): The CERN Library publishes its book catalog as Open Data. Pressemitteilung einsehbar unter <http://library.web.cern.ch/library/Library/announcement.html>.

Flimm, Oliver (2010): Offene Katalogdaten – Reaktionen. Einsehbar unter <http://blog.openbib.org/2010/03/18/offene-katalogdaten-reaktionen/>.

Free Software Foundation: The Free Software Definition. Einsehbar unter <http://www.gnu.org/philosophy/free-sw.html>.

Hochschulbibliothekszenrum des Landes Nordrhein-Westfalen (2010): Freigabe der Katalogdaten: Kölner Bibliotheken leisten Pionierarbeit. Pressemitteilung, einsehbar unter: <http://www.hbz-nrw.de/dokumentencenter/presse/pm/datenfreigabe>.

Library of Congress Working Group on the Future of Bibliographic Control (2007): Report on the Future of Bibliographic Control. Einsehbar unter <http://www.loc.gov/bibliographic-future/news/lcwg-report-draft-11-30-07-final.pdf>.

Möller, Erik (2005): Freiheit mit Fallstricken: Die Creative Commons NC-Lizenzen und ihre Folgen. Einsehbar unter <http://freedomdefined.org/Licenses/NC/De>.

Murray-Rust, Peter (2006): Open Data - the time has come. Einsehbar unter <http://wwwmm.ch.cam.ac.uk/blogs/murrayrust/?p=32>.

Open Source Initiative: Open Source Definition. Einsehbar unter <http://www.opensource.org/docs/osd>.

Open Knowledge Foundation (2007): Response to Working Group on the Future of Bibliographic Control (Library of Congress). Einsehbar unter <http://wiki.okfn.org/FutureOfBibliographicControl>.

Open Knowledge Foundation (2010): Entwurf eines Flyers zu Freien Katalogdaten. Die hier referenzierte Version 9 ist einsehbar unter <http://wiki.okfn.org/Entwurf%20eines%20Flyers%20zu%20Freien%20Katalogdaten?action=recall&rev=9>.

Open Knowledge Foundation: Open Definition. Einsehbar unter <http://www.opendefinition.org/okd/>.

Ostrowski, Felix / Pohl, Adrian (2010): Linked Data im hbz. Erscheint in B.I.T. Online 3/2010 und online u.a. unter <http://www.hbz-nrw.de/dokumentencenter/produkte/lod/>.

TED (2009): Tim Berners-Lee on the next Web. Einsehbar unter http://www.ted.com/talks/tim_berniers_lee_on_the_next_web.html.

TED (2010): Tim Berners-Lee: The year open data went worldwide. Einsehbar unter http://www.ted.com/talks/lang/eng/tim_berniers_lee_the_year_open_data_went_worldwide.html.

Wikipedia, Die freie Enzyklopädie (2010): Artikel „Crowdsourcing“. Bearbeitungsstand: 23. Juni 2010, 07:22 UTC. URL: <http://de.wikipedia.org/w/index.php?title=Crowdsourcing&oldid=75903904>.

Wikipedia, Die freie Enzyklopädie (2010): Artikel „Open Library“. Bearbeitungsstand: 9. Mai 2010, 02:20 UTC. URL: http://de.wikipedia.org/w/index.php?title=Open_Library&oldid=74131950.



Dieser Text ist unter folgende Creative-Commons-Lizenz veröffentlicht: [Creative Commons Namensnennung 3.0 Deutschland](https://creativecommons.org/licenses/by/3.0/de/).